Price 3d. Science Bulletin No. 22.

UNION OF SOUTH AFRICA

DEPARTMENT OF AGRICULTURE

AGRICULTURAL EXPERIMENT

ITS DESIGN AND INTERPRETATION

By E. PARISH, B.Sc.

Department of Agriculture

PRETORIA

THE GOVERNMENT PRINTING AND STATIONERY OFFICE

Printed by the Government Printer, Pretoria.

8272—6/12/21—1100

AGRICULTURAL EXPERIMENT: ITS DESIGN AND INTERPRETATION.

By E. Parish, B.Sc., Department of Agriculture.

Agricultural experiment, no matter how much care is taken in conducting it, is liable to error, due to random variation of the units composing it. This liability to individual variation has long been recognized and commented upon. Sir J. B. Lawes, in the course of a report on the comparative fattening qualities of sheep in an experiment, conducted at Rothamsted, some seventy years ago, in which 40 Hampshire and 40 Sussex wethers were under investigation, remarks:—

"It is perhaps seldom that animals have been drawn for purposes of experiment with more care than in the instances of which the foregoing tables record the results, yet we have scarcely a sheep in either breed which does not give twice, thrice, or more times as great an increase in gross live weight at one period as at another of equal length; whilst taking the entire period of the experiment, we have nearly double the increase with some animals as with others by their side and having ostensibly the same description and qualities of food rovided. . . .

The tenor of all published results on feeding seems to show that iese fluctuations and variations are the rule and not the exception; nd the fact of them, therefore, should lead us to great caution in rawing nice conclusions from experiments made with but a small umber of animals, and extending over a short period of time."

Similar extracts with reference to field experiments could be ıoted: instances of the magnitude of the variation in plots similarly ..eated apparently in all respects will be given later in this paper. Many cases could be cited with animal and crop experiments in which conflicting and opposing results have been obtained in successive years when experiments have been repeated.

Jevons many years ago wrote: "It is one of the most embarrassing things we can meet when experimental results agree too closely," and this statement was made concerning scientific experiment in which the conditions and factors are much more under control than in an agricultural experiment.

Absolute correctness in the results of agricultural trials is impossible to attain; the inevitable error may, however, be recognized and measured; and it is only by so doing that a true conclusion can be reached.

Unfortunately, it is still a common practice to plan experiments with single plots and single check or control plots, with the result that years of work and time and money are wasted and no really reliable conclusion reached within a reasonable time.

Still more unfortunately the practice of publishing results based on single plot trials continues, and the reader is either bewildered by the discrepancy of the results recorded in successive years or is deluded into accepting the unreliable result of a single year's trial.

An example may be given to illustrate the magnitude of the difference possible in the yields of plots similarly treated during any one season.

At Elsenburg in 1915 it was desired to commence a rotation experiment of extensive *versus* intensive farming, in order to test whether the often advocated intensive methods would really be profitable compared with the ordinary methods in practice in the district. In this experiment, owing to the areas being necessarily large in order that records of cost might be kept, it was not possible to replicate the plots to such an extent as to reduce the experimental error to a small percentage. It was therefore decided to treat all the eight plots required in the experiment similarly during the first season in order to ascertain the range of variation. The plots were each one acre in extent, and were arranged on the following plan. The yield in lb. of combined grain and straw are given, these figures being selected as involving less error than the weights of grain and straw separately.

1121	1178
1051	1224
935	981
400	767

The magnitude of the variation is too obvious to need comment. Perhaps the example may be a little unusual, but it furnishes a striking warning of the risk and liability to error involved in single plot trials, and of the necessity of so planning experiments that the error may be recognised and the extent of it measured.

The errors to which the results of field trials are subject, are of two kinds: (1) Those due to differences of soil and climate; (2) those chance, random, or casual deviations due to errors in weighing, uneven seeding and manuring and cultivation, effect of previous crops, birds, insects, etc. Errors of the first kind tend to be less in small plots than in large. Errors of the second kind tend to become less obvious as the size of the plot increases. It is found that little is to be gained in accuracy by increasing the size of the plot above about one-fortieth of an acre.

The Magnitude of Error due to Difference in Soil.

In a field trial the deviation due to soil can best be ascertained by an examination of the results obtained on the same plots over a number of years, since the variation of a random or casual kind tends to become evened out by such a method. The best figures available of the yields of crops grown on the same soil over a number of years are those obtained at the Rothamsted Experimental Station in England. At that station a manurial experiment on grass land has been carried on continuously from 1856 to date. Plots three and twelve were unmanured every year; the yields of these two plots furnish interesting statistical data of the variation in yield of the same plot, and of the difference in yield of the two plots, during this period. The plots were in a field as uniform as could ordinarily be obtained and the measurements and weights were recorded by skilled and experienced officers. Over the whole period of fifty years, one plot was found to yield on the average 10 per cent. higher than the other, and this difference must be attributed solely to the soil.

THE MAGNITUDE OF THE ERRORS OF A CASUAL KIND.

Errors of a casual or random nature due to all sorts of more or less accidental factors are liable to be overlooked, when experiments are being designed and interpreted. It is easy to understand that the soil in two plots may be different, and the yield of crops grown on them correspondingly different; but there is a tendency to regard the difference in yield as being constant and due only to the soil. Actually, however, the random deviation in any year may be so great as to outweigh any differences in soil. A reference to the figures obtained at Rothamsted quoted above, will make the magnitude of this difference apparent far better than any description possibly could. As stated above, over the whole period of 50 years the yield of the one plot averaged 10 per cent. better than that of the other. In odd years, however, the yield of the one plot was 49 per cent. greater, and 10 per cent. less than that of the other. These latter variations are of the random or casual kind.

Admitting now the existence of random variation, it might be supposed that this would in a period of, say, five years become evened out, and that a reliable conclusion might be derived on the average of a five years' experiment. In the experiment quoted above, the average of successive five-year periods of the yield of one plot compared with that of the other were:—

Periods.	Per cent.	Periods.	Per cent.
1856-60	105	1881-85	102
1861-65	121	1886-90	96
1866-70	128	1891-95	114
1871-75	121	1896-00	123
1876-80	108	1901-05	121

It is not so difficult to realize that there might be ups and downs of yield in odd years: it is much more difficult to anticipate that there could be so much divergence in the results of five-year averages.

No amount of care can wholly rid experiments of this random variation, but they may be so designed that the magnitude of it can be measured.

There are thus three matters to receive consideration in discussing the planning of field experiment:

1. The size of the plot.
2. The frequency and distribution of the plots.
3. The measurement of the error.

MEASUREMENT OF THE ERROR IN AGRICULTURAL EXPERIMENT.

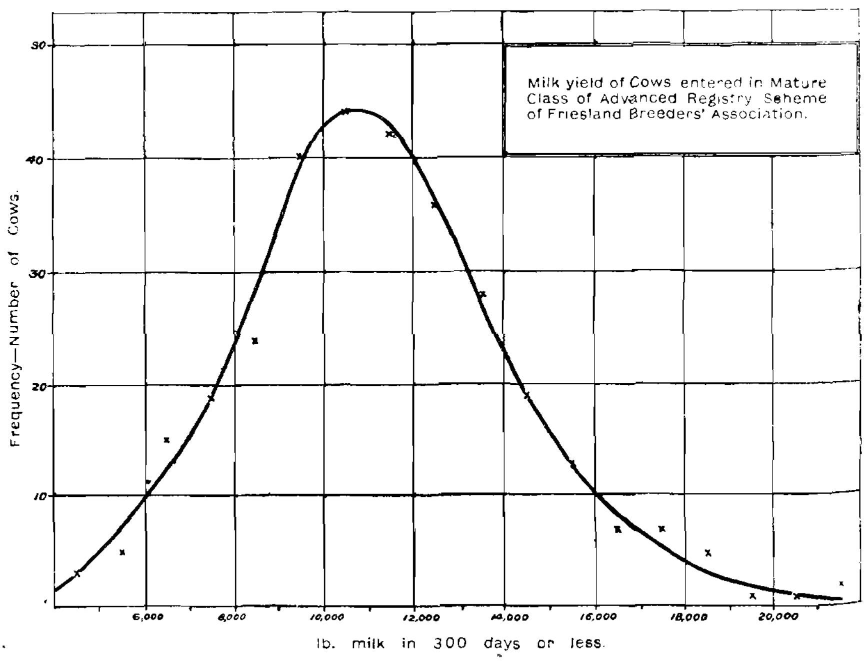
The method commonly employed with a group of results of the same nature or kind is to determine the average or arithmetic mean.

An average, however, is open to several objections; it may not correspond with the mode or normal result; it does not show the range of variation; and it can only be applied when the results are subject only to normal deviation, and are not disturbed by different factors. Before an average is used as a measure of or for representation of the results, they must be examined in order that it may be ascertained whether the variation is normal. This will most readily be discovered by means of a frequency curve. (It is obvious that if in any range of results two series are unwittingly included, then the average of the whole group can convey but little idea of the actual range of the individual.)

Records obtained in the Advanced Registry Scheme of the Friesland Breeders' Association of South Africa illustrate this point very clearly, and demonstrate also the use and value of a frequency curve. In this scheme the cows are divided, according to age, into five classes. Those over four years are all included in the mature class, and since this class contains more entries than the others, it is the most suitable for the purpose of illustration.

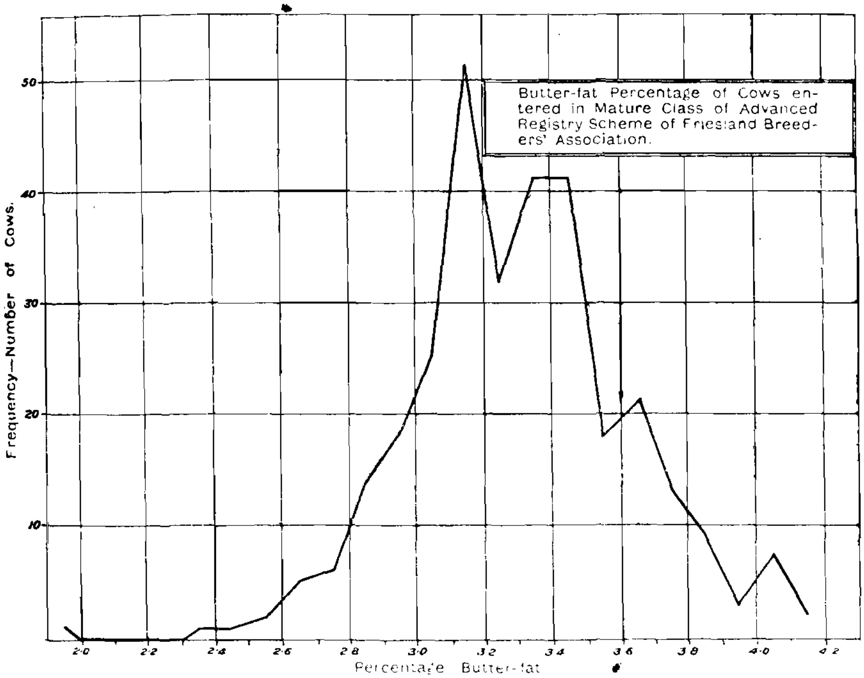
There are three sets of figures relative to the milk yield of each cow, viz. butter-fat percentage, total quantity of milk in the lactation, and total quantity of butter-fat. Since, however, the latter figure is a function of the other two, information to be gained from an examination of it, is not so easily obtained.

Below is plotted a frequency curve of the quantity of milk produced during the lactation, not including, however, any produced after 300 days in any case.



As will be seen the *fit* in this case is fairly good; the recorded results correspond as nearly as could be expected with the curve of normal variation. The class as a whole therefore is uniform. The greater number of the cows give between 10,000 and 12,000 pounds, and the mean yield is 11,340 lb. In such a case an average is of definite significance and value.

The curve of the plotted results of the percentage of butter-fat however, given below, is irregular and does not fit closely with a normal variation curve. There may be some disturbing factors producing the several maxima.



It may be noted that the percentage of butter-fat recorded for each cow is a figure compounded from the results of nine or ten monthly tests only, each of which is a composite of 4 to 6 tests. The total quantity of milk produced by each cow is a compound not of 40 or 50 records, but of two, and in some cases three records daily over the period of lactation, and thus the composite of about 600 to 900 records, numbers large enough to give a normal dispersion curve.

The arithmetic mean of the butter-fat percentage of the cows in this class is 3.302. As will be seen from the curve there are more cows giving 3.10 to 3.19 than there are 3.302. With such a group of results, exhibiting other than normal deviation, an average is a figure of but limited significance and value.

In order further to show the limitations of an average, the yields per morgen of maize given in the Census of Agricultural Producti n for 1918 may be used as an example. According to this Census, the average yield of maize in the three Provinces of Transvaal, Orange Free State and Natal, is 4.6 muids per morgen. If the districts be grouped according to yield, it is seen that the greatest number of districts, however, produce between 3.6 and 4.0 bags per morgen, while there are also more districts producing between 5.1 and 5.6 bags than there are producing 4.6, and the range of yield is from 1.82 to 11.67 bags per morgen. A bare statement of the average cannot adequately represent the range of yield, or show the yield obtained in the greatest number of districts.

Thus a figure is needed which will convey a more complete and accurate impression of the true position. This is provided by a figure expressing the degree of accuracy or reliability of any result and is known as the "probable error,"* this figure giving the range within which there is an even chance that any single result will fall. It has a mathematical basis, and is obtained by means of a formula $0.67\sqrt{\frac{\varepsilon d^2}{n-1}}$, where in a range of n results εd^2 is the sum of the squares of the differences from the mean. In a small number of results the probable error is itself liable to error.

The probable error of the mean in any range of results can also be determined by the formula $0.67\sqrt{\frac{\varepsilon d^2}{n(n-1)}}$. In this case also the probable error of the mean of a small number of results will not be so reliable as that of a large.

It is clear that, if the probable error in terms of the mean in any range of results is high, then correspondingly little reliance can be placed on any one result or much significance be attached to the mean. The calculated odds against any one result differing from the average by more than twice the probable error is 4½ to 1, and the odds against any one result differing from the average by more than three times the probable error is 22 to 1, while a 30 to 1 odds, which Professor Wood regards as decisive enough for all practical purposes, is obtained when the difference is 3.2 times the probable error, it being always borne in mind that the probable error figure itself is not so reliable for a small range at figures as for a large.

An example may be given to illustrate the application of this method to the results of experiments.

In 1915 an experiment was commenced at Elsenburg to test the effect on a subsequent cereal crop of fallowing or "braak" against a green manuring. In 1915 the plots were variously green manured and fallowed or "braak" and in 1916 an oat crop was planted on all plots. In 1917 and 1918 a similar succession was conducted but with wheat as the cereal instead of oats. The plots were each one-tenth

---

* A figure easier to calculate and generally more useful is the "standard deviation" obtained by the formula $\sqrt{\frac{\varepsilon d^2}{n-1}}$, for a single result. The probable error has been used in this paper rather than the standard deviation, since it has a more practical significance, and is easier of conception.

acre, and the arrangement, and the number of them, with th weights of cereal in lb. of grain per plot, are shown below:—

Treatment, 1915 and 1917.	Yield, Oats, 1916.	Yield, Wheat, 1918.
Fallow	171	118
Green Manure	161	75
Fallow	172	113
Green Manure	135	89
Fallow	189	127
Green Manure	186	87
Fallow	231	113
Green Manure	207	111
Fallow	203	124
Green Manure	198	113

The mean weights of grain per one-tenth acre plot were as follows:—

	Oats, 1916.	Wheat, 1918.
Fallowed plots...	193	119
Green manured plots	177	95
Average increase due to fallowing ...	16	24
Percentage increase	8.3%	20.1%

Now the weights of grain harvested from the different plots in each series vary greatly, and it is essential to know what reliance can be placed on the mean yields of 193 and 177 lb. in 1916, and 119 and 95 in 1918, and on the mean differences of 16 and 24 lb. respectively in favour of the fallowing.

The probable errors of the means using the formulae already given, are found to be:—

	Oats, 1916.	Wheat, 1918.
Fallowed plots	$193 \pm 7 \cdot 5$	$119 \pm 1 \cdot 9$
Green manured plots	$177 \pm 8 \cdot 7$	$95 \pm 4 \cdot 9$

These results are more easily comparable if given as percentages of the mean as follows:—

	Oats, 1916.	Wheat, 1918.
Fallowed plots	3.9%	1.6%
Green manured	4.9%	5.1%

When the results are presented in this way, it is apparent that a difference in the neighbourhood of 5 per cent. is of little significance since the averages themselves have a probable error between from 1.6 to 5.1 per cent.

Fortunately, a formula is available whereby the probable error of the difference and thus the exact significance of this difference can be determined. The probable error of the difference of two probable errors E_1 and E_2 is given by $\sqrt{E_1^2 + E_2^2}$. Applying this formula, the probable error of the average difference of 16 lb. per plot between the fallowed and green manured plots in 1916 is 11.47 lb., and the corresponding figure for 1918, with a difference of 24 lb. is 5.25 lb.

The definition of probable error, as given above, was that in a range of results the chances are even that any one result will lie within the range of the probable error from the mean. It is apparent therefore that if a difference between two sets of results is of the same order of magnitude as the probable error of the difference of these results, such difference can be of little significance. It is apparent also that according as this difference increases proportionately to its probable error so will the odds against such difference being due to normal variation increase. The following table gives the odds corresponding to specified differences:—

Difference from mean in terms of Probable Error.	Odds against such difference being due to normal variation.
1.00	1 to 1
1.25	3 to 2
1.44	2 to 1
1.71	3 to 1
1.90	4 to 1
2.00	9 to 2
2.50	10 to 1
2.93	20 to 1
3.00	22 to 1
3.20	30 to 1
4.00	140 to 1
4.90	1000 to 1
5.00	1350 to 1

Considering again the experiments referred to above, it is noted that the mean percentage difference of 8.3 per cent. obtained in 1916 is not much larger than the probable error of each mean and therefore is not decisive. The difference of 20.1 per cent. obtained in 1918 is, however, approximately 4 times its probable error and therefore is of a decisive nature. The odds against such difference being due to normal variation are 140 to 1. Whether this greater difference obtained in 1918 than in 1916 is due to cumulative effect is not clear: the yields of winter cereals in general in 1918 were less than in 1916, and naturally in a bad year, with a deficient rainfall, fallowing might be expected to produce a proportionately greater benefit than when the rainfall were less deficient.

A further case may be quoted in order to show the difficulty of correctly interpreting the results of agricultural experiment.

At Elsenburg in 1916 an experiment was commenced with the object of investigating the method by which the grain rotation of the Western Province of the Cape could best be improved, whether by ploughing under a green crop of oats and vetches, or by feeding it off, or by making into hay. It was expected that ploughing under a green crop would result in the greatest yield of the subsequent cereal crop, but it was obviously possible that such increased yield might be obtained at such cost as to be unprofitable. Unless the benefit derived by each method could be determined fairly accurately, the experiment would provide little information either novel or useful. The experiment was planned as follows: this arrangement of plots permitting of the measurement of the experimental error, and of the elimination of any difference due to progressive change in soil. The plots were one-tenth of an acre in extent; the yields of wheat grain in lb. per acre are shown.

Yields are given in lb. of grain per acre:—

No. of Plot.	1916 Vetches and Oats.	Yield of Wheat, 1917.	Yield of Wheat, 1918.
15	Made into Hay ...	465	320
14	Ploughed under ...	700	370
13	Fed off by Sheep ...	630	490
12	Fed off by Sheep ...	580	450
11	Ploughed under ...	*715	420
10	Made into Hay ...	440	350
9	Made into Hay ...	300	320
8	Ploughed under ...	*715	360
7	Fed off by Sheep ...	350	340
6	Fed off by Sheep ...	550	400
5	Ploughed under ...	1235	430
4	Made into Hay ...	630	520
3	Made into Hay .	490	400
2	Ploughed under ...	1035	450
1	Fed off by Sheep ...	505	540

* Owing to choking of the thrasher, yield of plots 11 and 8 in 1917 were combined: in the calculations made below, half of 1430 has been credited to Plot 8 and half to Plot 11.

The average weights of grain in lb. per plot in each series, with the probable errors of the averages, obtained by calculation on the formula given above are as follows:—

Treatment, 1916.	Average weight of grain. 1917.	1918.
Vetch and oat crop fed off	523±32	444±23
Vetch and oat crop ploughed under...	880±73	406±12
Vetch and oat crop made into hay...	471±33	382±25

In terms of the mean yield in each case, the probable errors of these averages are:—

Treatment, 1916.	1917.	1918.
Vetch and oat crop fed off...	6.1%	5.1%
Vetch and oat crop ploughed under...	8.3%	2.7%
Vetch and oat crop made into hay...	7.0%	6.5%

In 1917 the ploughed-under plots gave 257 lb. per acre, equivalent to 29.2 per cent., more than the fed-off plots, and 409 lb., equivalent to 46.5 per cent., more than the made-into-hay plots. These differences are respectively two and five times the probable error and are therefore fairly decisive.

The average of the fed-off plots in 1917 was 52 lb. per acre (equivalent to 9.9 per cent.) greater than the average of the hayed plots. This difference is of the same order of magnitude as the probable error and therefore is not significant.

In 1918 the average of the yield of wheat from the plots which had been fed off in 1916 was 38 lb. per acre, or 8.6 per cent. greater than from the ploughed-under plots, and 62 lb., or 14 per cent. greater than from the hayed plots. The first difference is of the same order of magnitude as the probable error and is therefore insignificant. The second figure is just over twice the probable error and is therefore more significant, though hardly decisive.

This experiment furnishes interesting instances of the variation in the yield of contiguous plots treated similarly, and provides a grave warning to those experimenters who use but one control or check plot, and then try to argue on differences in the region of 5 and 10 per cent.

In 1917, from contiguous plots treated as far as possible identically in 1916 and 1917, the yields were 58 and 63 lb., with a difference of 9 per cent.; 33 and 44 lb., with a difference of 33 per cent.; 55 and 35 lb., with a difference of 57 per cent.; and 49 and 63 lb., with a difference of 28 per cent. This difference in the yield of adjacent plots treated similarly in all respects, of from 9 per cent. to 57 per cent., is surely striking, and should be borne in mind by all experimentalists in planning and interpreting their experiments.

Probable Error at Elsenburg.

From these experiments the probable error of one plot at Elsenburg appears to be about 11 per cent., and of the average of 5 plots 5 per cent. The results obtained in subsequent years, permitting of calculation from a larger number of experiments, may, of course, make it necessary to modify this figure to some extent. On this basis therefore, differences in the yield in single plot trials, unless exceeding 25 per cent., are not decisive, but can only be regarded as indicative; and in the average of five plots, differences not exceeding

That the conditions in the Elsenburg experiment were not exceptional, is shown by results obtained at Potchefstroom. In a manurial rotation experiment conducted there since 1918, are 16 plots in each of four ranges. Plots 3, 7, 11, and 15 in each range are check plots and are treated similarly. The probable error of a single result expressed as percentage of the mean of the four check plots, with maize as the crop, during three years has been variously 20.4 per cent., 19.2 per cent., 34.2 per cent. and 14.9 per cent. These results are somewhat high, and in this experiment an increase or decrease due to the manure under trial would have to be very high indeed before being decisive. It is recognized that the probable error itself is liable to error when calculated over such a small number of results; nevertheless, the individual results (e.g. in 1919 the yield of the check plots Nos. 3, 7, 11. and 15 were respectively 60.5, 97, 73, and 47 lb. per plot of one-twentieth acre) are so divergent as to indicate very plainly that it will be many years before any conclusive results can be obtained from this experiment while conducted under present conditions.

On the other hand, an investigation conducted at Cedara in 1920, shows that under certain conditions the probable error of a single plot may be brought within a reasonable figure. A ten acre field of maize was planted and divided carefully into ten one-acre plots. These plots were harvested separately and carefully and the probable error of the yield of grain from a single plot expressed in terms of the mean was 3.98 per cent., and the probable error of the ears similarly expressed was 3.88 per cent., thus showing fairly good correlation. A similar experiment on a more extensive scale was commenced in 1920-1921 with 100 plots of maize, but a hailstorm in March last unfortunately wiped out the crop, thus preventing any results being obtained.

Size of Plots and Amount of Replication.

A large number of tests, made in various countries, has established the fact that the results of single plot trials, whether these plots be large or small, are unreliable and subject to a probable error of from 5 to 10 per cent. Increase in the size of the plot tends to reduce the random variation due to irregular stand, inaccuracies in weighing, individual variation of the plants, attacks of birds, insects, etc., but this reduction is counterbalanced, and may be outbalanced, by differences of soil, which in general, tend to be greater in large areas than in small. The random or casual variations cannot be eliminated, but by systematic replication of the plots, they can be reduced and the limitations of the result be determined. Since this replication is essential, it becomes obvious that large size in plots is mere waste of ground and labour. Moreover, with large plots an error may be introduced due to the weather if it be not possible to get any set of cultural and harvesting operations performed in a day.

On the question of the size of plot there appears to be some difference of opinion among authorities. Hall, of the Rothamsted Experimental Station, concluded in 1911 that little advantage is to be gained by increasing the size of the plot above one-fortieth of an acre. He also considered that for practical purposes in any field experiment, each unit of comparison should be given five plots, or in other words be replicated five times, these plots being systematically distributed within the experimental area. By this method the two per cent., and these conclusions and recommendations of Sir A. D. Hall have never since been contested.

There remain, however, several questions of secondary importance for consideration.

The arrangement in the replication must not only be systematic, but must be such as to vitiate progressive or systematic differences in the soil. For instance, if the experiment be laid down the slope of a hill as below:—

A
B
C
A
B
C
A
B
C
A
B
C

A systematic and possibly constant error might be introduced which would vitiate the comparative result of every three plots and thus of the whole.

The best arrangement in this experiment would be:—

A
B
C
C
B
A
A
B
C
C
B

This system is suitable for manurial, or rotation or other trials where the effect of one or two factors is required to be determined.

Experience has shown that it is advisable to limit the factors—of which the effect is under test—in any experiment to two or three, otherwise the plots in each series, and the replications in the different series may be so far removed from each other as to give results into which large errors may have been introduced. If the effect of several factors is to be determined, these should be broken up and tested in a series of experiments each with check or standard plots, rather than combined in one gigantic trial.

Where it may be necessary to test a number of different factors in one experiment, the plots may be arranged on the system devised by Dr. Sonne, the Danish Experimenter, whereby there are five plots of each kind scattered systematically about the field.

A	F	E	G
B	G	D	F
C	A	C	E
D	B	B	D
E	C	A	C
F	D	G	B
G	E	F	A

The possible differential effect of the border on the crop in the different plots must be borne in mind and guarded against. Owing to the lack of competition along the alley side of the border, the marginal rows usually grow better than those in the centre of the plot. The extent of the errors introduced from this cause were found by Arny and Hayes in the United States in plots 8.25 feet wide, to vary from 7.48 to 15.78 per cent., with an average of 12.78 per cent. for 11 varieties of oats, and for 5 varieties of wheat to vary from 14.07 to 23.51 per cent.

This error may be guarded against by planting one or preferably two rows of the same variety beyond the margin of the plot, and then cutting out these rows at harvesting. This method is better than arranging the different plots without any alleys, inasmuch as the crop should be tested in competition with itself rather than in competition with other crops. According to Kiesselbach of the Nebraska Agricultural Experiment Station, the competition of neighbours may have a pronounced effect upon the yield

Variety Trials.

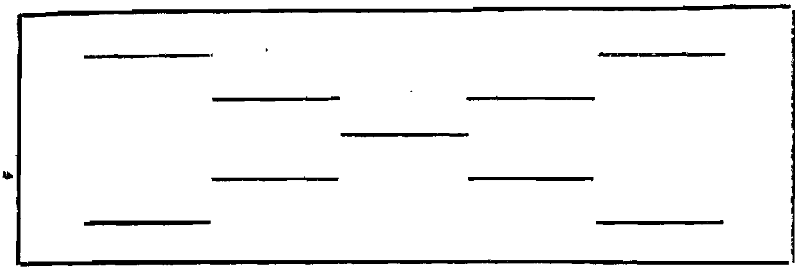
From what has been said above, it is obvious that it will be advisable to arrange the plots as narrow as can conveniently be worked. In manurial trials there are the difficulties of manuring and, if the trial be permanent, of preventing the surface soil being ploughed beyond the limits of the plot; these difficulties, however are not encountered in variety trials.

Single row trials are to be avoided where possible owing to the lack of competition, or competition with other varieties, discussed above. Plots should be at least 5 rows wide, and preferably 9 or 10, in order to allow of the border rows being cut out. The best arrangement seems to be that suggested by Spragg in his paper on the co-efficient of yield in No. 5, Vol. 12, of the *Journal of the American Society of Agronomy*. He suggests that owing to the effect of climate on the yield, in each variety trial one check plot of a standard high yielding variety should be planted to each two or four varieties, and that the yield then be given not in bushels per acre, but in terms of the yield of the standard variety in the check plots. By this method a *coefficient of yield* is calculated which furnishes a year-to-year standard for comparison. Spragg also considers that the plots should be long and narrow, and that the distance between the checks should not be more than one-twentieth of the length of the plots. By this system of arranging the plots long and narrow, and of stating all yields in terms of the yield of the standard variety in the check plots, Spragg considers that the necessity for replication in order to reduce the experimental error is obviated, and that the long plot can be harvested as a unit, and need not be split up.

In ear-to-row testing of maize, according to Lyman Carrier of the United States Department of Agriculture, errors may be introduced, inasmuch as cross-pollination may influence the size and yield of the grain the same season that the cross is made.

Co-operative Experiment.

In co-operative experiment the difficulty of harvesting carefully areas of from one tenth acre to one acre or more, frequently arises, and may in some cases be insurmountable. Some recent work by Arny and Garber and by Arny and Steinmetz in the United States indicates a way out of this difficulty. These observers found that by harvesting carefully a number of accurately measured small areas distributed systematically over a plot, a result may be obtained which gives as accurate a result as harvesting the whole area. The units on which their work was based consisted of one row 5½ yards long, and one square yard respectively, and the number of these required to give a result as reliable as that obtained from harvesting the whole area, were found to be from 9 upwards in the rod-row method, and from 5 to 10 in the square yard method, according to the uniformity of the stand. It is not considered necessary to harvest the 5 or 9 small units separately: what is important is the systematic distribution. This must not be random or it may be subject to bias, but must be on a definite system. In the rod-row method the distribution proposed by Arny and Garber in the plot is:—



Marking of Experiment Plots.

The method adopted at Rothamsted may confidently be recommended for adoption in all field trials, whether temporary, permanent, or co-operative. At this station the plots are marked out initially, care being taken to keep the boundaries of the plots well away from trees, roads, or other disturbing factors. The outside lines of the experimental area are then marked out by fairly stout permanent posts set in the fences and thus out of the way of implements and animals. At the corners of the actual plots, creosoted stakes 1 foot long by 2½ inches square are driven in until the top is one foot below the surface. At any time when the limits of the plots are required to be known, the outside fence posts are used as sighting lines, measurements are taken along these lines, and the ground probed in the expected position of the posts until they are found.

Probable Error in Feeding Trials.

Fattening Experiments.—The probable error in live weight increase of one animal in feeding trials with cattle and sheep is 14 per cent., according to Professor T. B. Wood, this figure being obtained by examination of 23 experiments conducted in England and America with a total of 230 cattle and sheep.

Mitchell and Grindley, of the Illinois Experiment Station, consider this figure is correct for sheep but find from their investigation that 11 per cent. is the correct figure for steers. The probable error of one animal in the feeding trials with cattle conducted at Potchefstroom and Cedara for the years 1914 to 1919, taking young and old, is 18.6 per cent. In these experiments 95 animals were used with an average of 4.5 animals per group, whereas in the experiments quoted above the average per group was 12 animals, giving therefore a more reliable figure.

With a probable error of 18.6 per cent. for a single animal, the probable error of the average of a group of 4 animals would be 9.3 per cent. In the South African experiments therefore a difference between the average of one group and another if less than 35 per cent. could not be considered as decisive and conclusive, but only as indicating a "probable" difference, the measure of certainty of the result naturally increasing according to the amount by which the difference exceeded the probable error.

In these South African feeding trials the probable error in the experiments with old oxen was very high. Omitting the results with the old oxen, the probable error of a single animal from the 68 young animals in 15 groups was 10.9 per cent., giving a probable error of the average of a group of four as 5.4 per cent., a more reasonable figure.

Sufficient data regarding experiments with pigs in South Africa have not yet been published to allow of the calculation of the probable error for these animals. An inquiry made by Robinson and Halnan, using results of experiments in Scotland, England, and the United States, revealed the probable error of one animal to be in the region of 10 per cent. of the average live weight increase. As a result of this inquiry 12 weeks appeared to be the shortest period during which a trial should be carried on with pigs, otherwise the probable error were likely to be unduly high.

Crowther, of Leeds, from an examination of various pig-feeding experiments, concluded that individual feeding resulted in less variation within a group than collective feeding.

Mitchell and Grindley, in Bulletin No. 165 of the Illinois Agricultural Experiment Station, from their somewhat extensive investigations conclude:—

1. Such experimental conditions as are favourable to growth and fattening are favourable to uniformity of individual gains. As corollaries to this may be stated: given two groups of animals under different conditions, that group growing or fattening at the more rapid rate, will, in general, tend to exhibit the more uniform gains, uniformity being measured on the percentage scale. If change in ration, weather, or other conditions is resulting continually in conditions more favourable to growth or fattening, the gains within the group will tend to become more and more uniform. From this it follows that a constant ration throughout the experiment results in progressively less favourable conditions.

2. Usually the gains at the beginning of an experiment are extremely variable.

3. Effect of changes in experimental conditions on the variability of gains is more favourable and more noticeable in the case of sheep than in the case of either steers or swine.

4. Great care in selection of animals for a feeding experiment is very important: the animals in the lots under comparison should be of the same breed and type, sex, age, and, as far as possible, should have been under the same treatment for some time previous.

5. No benefit is to be gained by putting the animals under test before the experiment starts, and selecting only those which show uniform gains during this preliminary period.

6. The lots of animals employed should be fairly large, 10 to 15 animals per lot being considered to be the minimum. Large lots of animals, however, should not be the excuse for poor selection.

7. The feeding stuffs used in a feeding experiment should be analysed, since owing to the variation in the composition of feeding stuffs, the adoption of average analyses previously determined is liable to lead to error.

8. Experiments should in most cases be repeated before definite conclusions are published.

Feeding Experiments for Milk Production.

These present peculiar difficulty, inasmuch as not only is there a possibility of gain or decrease in live weight to be contended with, but also that the milk yield of cows is subject normally to a good deal of variation both of the random and constant kinds. It is affected by the weather, by slight indisposition, by oestrum, by pregnancy, by the time of calving, by change of attendant, and varies with the stage of the lactation. To the numerous possible disturbing factors to be met in fattening experiments is added the much more complex and delicate function of milk secretion. Not only are the selection of the animals and the control of the conditions more difficult in milk production experiments than in fattening, but the interpretation of the results is even more perplexing. The progressive change with the stage in the lactation period is one of the greatest difficulties that has to be faced, for it tends to prevent the experiments being carried on over sufficiently long periods to allow of decisive results being obtained. Both the continuous and alternation systems, and a combination of the two have been tried, and faults have been found with all three methods. Maynard and Myers, in Bulletin No. 397 of the Cornell University Experiment Station, state: "The advantage of the alternation system lies in the fact that the trial of the two sets of rations with the same animals eliminates many of the factors due to individuality and obviates uncertainties in attempts to make up groups of like production. On the other hand, the use of the alternation system necessitates a short feeding period." This is a disadvantage in that not only is the change itself disturbing in its effect, but that foods may have both a delayed and residual effect—it is to be expected that the effect of a certain food on the milk production will lag behind the actual feeding of it. There is the further difficulty that if the milk yield of an animal is allowed to drop through faulty feeding or insufficient nutrition, difficulty will be experienced in getting the yield back to normal, and this difficulty may in some cases be insurmountable. If the alternation system is used the feeding periods should be at least four weeks and preferably more. Periods of two or three weeks are too short.

The difficulty in the continuous system is well voiced by Hills*: "He who can from an often limited number of animals formulate groups which for a great length of time will prove essentially equivalent in their milk-making powers is gifted with second sight."

Maynard and Myers in Bulletin 397 of the Cornell University Experiment Station give the following summary concerning the selection of cows for feeding experiments in milk production.

1. Groups should be selected on the basis of production for a trial period just preceding the experimental period.

2. The cows in a group should be of approximately equal individual production: a range of 100 lb. milk in two weeks' yield is permissible.

3. The cows in a group should be as nearly as possible of the same age. There is no indication that cows of any particular age are more variable than those of any other age, but for other reasons it seems desirable to exclude heifers.

---

* *Bull. 181, U.S. Department of Agriculture*, 1907.

4. Cows in a group should be in approximately the same stage of lactation. There is apparently no increase in variability caused by mixing cows in different stages of lactation, so long as the experimental period is completed before the rapid decrease in production begins. According to Woll (1912), this rapid decrease begins at about the eighth month, and hence, if a period of about five months is to be used, the cows should have freshened not more than three months before the beginning of the period.

5. Only healthy, vigorous, normal, mature cows, should be included.

6. Groups should be made as large as possible without sacrificing more important factors.

Error due to Bias in the Observer.

The bias referred to is of the involuntary kind and cannot wholly be eliminated, though by care and training it may be reduced. An instructive example of the possibility of error of this type is afforded by plotting the frequency curves of the results recorded in the United States Department of Agriculture Year-book for 1918, of 169 observations of the depth of spring and fall ploughing in Indiana. This curve shows maxima at 5, 6, 7, and 8 inches and minima at the half inches. It is extremely improbable that such a curve accurately represents the actual depths of ploughing; it does, however, represent the *recorded* depths of ploughing. In this case the error is not in the results but in the observers, and is due to a tendency in the majority of them to observe and record the depths in units of inches rather than in halves. From these recorded results a corrected curve could of course be drawn which would fairly well represent the true depths, but that is outside the province of these remarks.

Summarized Suggestions on the Conduct of Agricultural Experiments.

1. The variation in yield of single plots similarly treated is so great as to render valueless comparisons made from the result of single plot trials; similarly with experiments with small numbers of animals, owing to the high variation in the individual.

2. Replication in agricultural experiments is absolutely necessary. This should be obtained rather by replication of the plots in a field trial, or animals in a feeding trial, in any year, than by continuing the experiments over a number of years. It is desirable, however, even when the experiments are properly designed, that they be repeated in successive years.

3. In field trials the replication of the plots must be systematic and so designed as to vitiate the effect of progressive differences in the soil. Suggestions for the conduct of manurial and variety trials and co-operative experiments are made in the body of this paper.

4. In animal feeding trials the individuals must be carefully selected and be uniform in age, breed, sex, and condition, and should be at least five in number in each lot, and preferably ten. The animals in the lots under comparison should be subjected to uniform conditions for two or three weeks prior to the commencement of the experiment.

5. In field trials no appreciable increase in reliability is gained by enlarging the plots beyond one-fortieth of an acre.

LITERATURE AND REFERENCES.

"The Experimental Error in Field Trials." Mercer, W. B., and Hall, A. D., *Jour. Agric. Sc.*, Vol. IV, No. 2 (1911).

"The Interpretation of Experimental Results." Wood, T. B., and Stratton, F. J. M., *Jour. Agric. Sc.*, Vol. III, No. 4 (1910).

"Field Technic in Determining Yields of Plots of Grain by the Rod-Row Method." Arny and Garber, *Jour. Amer. Soc. Agron.*, Vol. II, No. 1, pp. 33-48.

"Field Technic in Determining Yields of Experimental Plots by the Square Yard Method." Arny and Steinmetz, *Jour. Amer. Soc. Agron.*, Vol. II, No. 3, pp. 81-106.

"A Reason for the Contradictory Results in Corn Experiments." Lyman Carrier, *Jour. Amer. Soc. Agron.*, Vol. II, No. 3, pp. 106-114.

"Experimental Error in Field Trials." T. A. Kiesselbach, *Jour. Amer. Soc. Agron.*, Vol. II, No. 6, pp. 235-242.

"Plat Competition as a Source of Error in Crop Tests." T. A. Kiesselbach, *Jour. Amer. Soc. Agron.*, Vol. II, No. 6, pp. 242-247.

"The Unreliability of Short Time Experiments." Harris and Butt, *Jour. Amer. Soc. Agron.*, Vol. 12, No. 5.

"The Coefficiency of Yield." Frank A. Spragg, *Jour. Amer. Soc. Agron.*, Vol. 12, No. 5, pp. 168-175.

"Experiments in Field Technic in Plot Tests." Arny and Hayes, *Jour. Agric. Res. U.S.A. Dept. Agric.*, Vol. 15, No. 4.

"The Element of Uncertainty in the Interpretation of Feeding Experiments." Mitchell and Grindley, Bull. 165, Illinois Experiment Station.

"The Probable Error of a Mean." *Biometrika*, Vol. 6, No. 1.

"The Refinement of Feeding Experiments for Milk Production by the application of Statistical Methods." Maynard and Myers, Bull. 397, Cornell University Experiment Station.

"An Introduction to the Theory of Statistics." Udney Yule, Griffin & Co.

"Davenport's Statistical Methods."